

## Transcriptomic silencing as a potential mechanism of treatment resistance

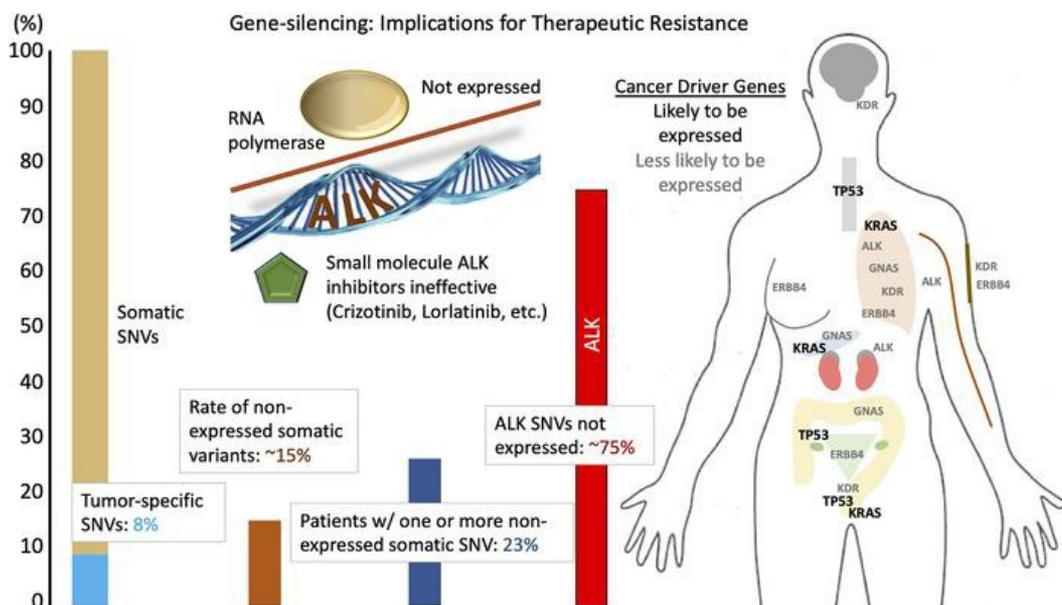
Jacob J. Adashek, ... , Sandeep K. Reddy, Razelle Kurzrock

*JCI Insight.* 2020;5(11):e134824. <https://doi.org/10.1172/jci.insight.134824>.

Research Article

Oncology

### Graphical abstract



Find the latest version:

<https://jci.me/134824/pdf>



# Transcriptomic silencing as a potential mechanism of treatment resistance

Jacob J. Adashek,<sup>1</sup> Shumei Kato,<sup>2</sup> Rahul Parulkar,<sup>3</sup> Christopher W. Szeto,<sup>3</sup> J. Zachary Sanborn,<sup>3</sup> Charles J. Vaske,<sup>3</sup> Stephen C. Benz,<sup>3</sup> Sandeep K. Reddy,<sup>4</sup> and Razelle Kurzrock<sup>2</sup>

<sup>1</sup>Department of Internal Medicine, University of South Florida, H. Lee Moffitt Cancer Center & Research Institute, Tampa, Florida, USA. <sup>2</sup>Center for Personalized Cancer Therapy and Division of Hematology and Oncology, Department of Medicine, University of California, San Diego, Moores Cancer Center, La Jolla, California, USA. <sup>3</sup>NantOmics, Santa Cruz, California, USA. <sup>4</sup>NantHealth, Culver City, California, USA.

Next-generation sequencing (NGS) has not revealed all the mechanisms underlying resistance to genomically matched drugs. Here, we performed in 1417 tumors whole-exome tumor (somatic)/normal (germline) NGS and whole-transcriptome sequencing, the latter focusing on a clinically oriented 50-gene panel in order to examine transcriptomic silencing of putative driver alterations. In this large-scale study, approximately 13% of the somatic single nucleotide variants (SNVs) were unexpectedly not expressed as RNA; 23% of patients had  $\geq 1$  nonexpressed SNV. SNV-bearing genes consistently transcribed were *TP53*, *PIK3CA*, and *KRAS*; those with lower transcription rates were *ALK*, *CSF1R*, *ERBB4*, *FLT3*, *GNAS*, *HNF1A*, *KDR*, *PDGFR*, *RET*, and *SMD*. We also determined the frequency of tumor mutations being germline, rather than somatic, in these and an additional 462 tumors with tumor/normal exomes; 33.8% of germline SNVs within the gene panel were rare (not found after filtering through variant information domains) and at risk of being falsely reported as somatic. Both the frequency of silenced variant transcription and the risk of falsely identifying germline mutations as somatic/tumor related are important phenomena. Therefore, transcriptomics is a critical adjunct to genomics when interrogating patient tumors for actionable alterations, because, without expression of the target aberrations, there will likely be therapeutic resistance.

**Authorship note:** JJA and SK contributed equally to this work.

**Conflict of interest:** SK serves as a consultant for Foundation Medicine. RP, CWS, JZS, CJV, and SCB are employees of NantOmics. SKR is an employee of NantHealth. RK has research funding from Incyte, Genentech, Merck Serono, Pfizer, Sequenom, Foundation Medicine, Guardant Health, Grifols, and Konica Minolta as well as consultant fees from LOXO, XBiotech, Actuate Therapeutics, Genentech, and NeoMed. RK receives speaker fees from Roche and has an equity interest in IDbyDNA and Curematch.

**Copyright:** © 2020, American Society for Clinical Investigation.

**Submitted:** November 6, 2019

**Accepted:** April 29, 2020

**Published:** June 4, 2020.

**Reference information:** *JCI Insight*. 2020;5(11):e134824.  
<https://doi.org/10.1172/jci.insight.134824>.

## Introduction

Clinical-grade next-generation sequencing (NGS) is being increasingly used as a tool to direct the treatment of cancers. By revealing alterations in some of the most commonly affected and instrumental genes in cancer development and progression, NGS ostensibly identifies potentially actionable targets (1). The benefit of NGS has been recognized by the Centers for Medicare & Medicaid Services, which led to their expansion of NGS coverage for patients with advanced cancers (2). The information obtained from interrogating a single patient's cancer with NGS can potentially help guide his or her care, while the data derived from population-based NGS studies can potentially shed light on previously unknown drivers of response and resistance across cancer subtypes (3).

Unfortunately, the outcome for many patients with metastatic tumors remains grim even in the era of precision oncology, with responses often being unpredictable, even when clinical-grade NGS is used. As tumors are exposed to successive lines of therapy (e.g., chemotherapy, radiation therapy, gene- or immune-targeted agents, etc.), tumor heterogeneity evolves and new drivers may emerge (4). Additional resistance mechanisms may be due to one or more of the following: the tumor microenvironment, which secretes various hormones, growth factors, and cytokines that can propel cell growth and survival (5); cancer progenitor cells that increase the expression of drug efflux transporter genes that pump drugs out of the cell and into the interstitium (6); mutations within drug transporter genes that attenuate drug binding and influx into the cell and/or absorption (7); increased expression of antiapoptotic proteins and decreased expression of proapoptotic proteins (7); amplification of oncogenes and protumorigenic genes (8); mutations within drug target domains, such as the kinase motifs (9); and changes in genes, such as *BRCA*, that revert them to “normal,” consequently abrogating sensitivity to PARP inhibitors (10). Altered expression, including gene silencing, via mechanisms, such as epigenetic changes (DNA methylation and histone

acetylation) (11), miRNA (12) and/or alternative splicing of RNA and/or fluctuations in RNA-binding proteins (13, 14), may also play a role in resistance to therapy. However, there is a dearth of knowledge about the frequency of gene silencing in cancer tissue and its potential implications for drug resistance.

Transcriptomic analysis can be used to reveal the silencing of cancer-related variant transcription. Others have advanced the use of combined RNA and DNA analysis, including the development of tools such as RADIA (15) and UNCeQR (16), with the primary goal of enhancing detection of somatic mutations, especially at low DNA allelic frequencies. These studies and those focused on RNA analysis (17, 18) have revealed the importance of looking at expression for understanding tumor biology and cancer progression. The work presented here builds on the aforementioned studies; however, it was performed using a very large data set with the goal of determining the frequency of silencing of single nucleotide variant-bearing (SNV-bearing) genes (that are potentially drivers) by integrated RNA and DNA analysis.

Herein, we examined a cohort of 1879 patients with whole-exome sequencing of tumor and nontumor tissue (tumor/normal sequencing) to determine the somatic versus germline origin of identified mutations, in concert with cognate transcriptomic sequencing in a subset of 1417 of these patients, in order to evaluate the frequency of tumor-related gene silencing in the pan-cancer setting.

## Results

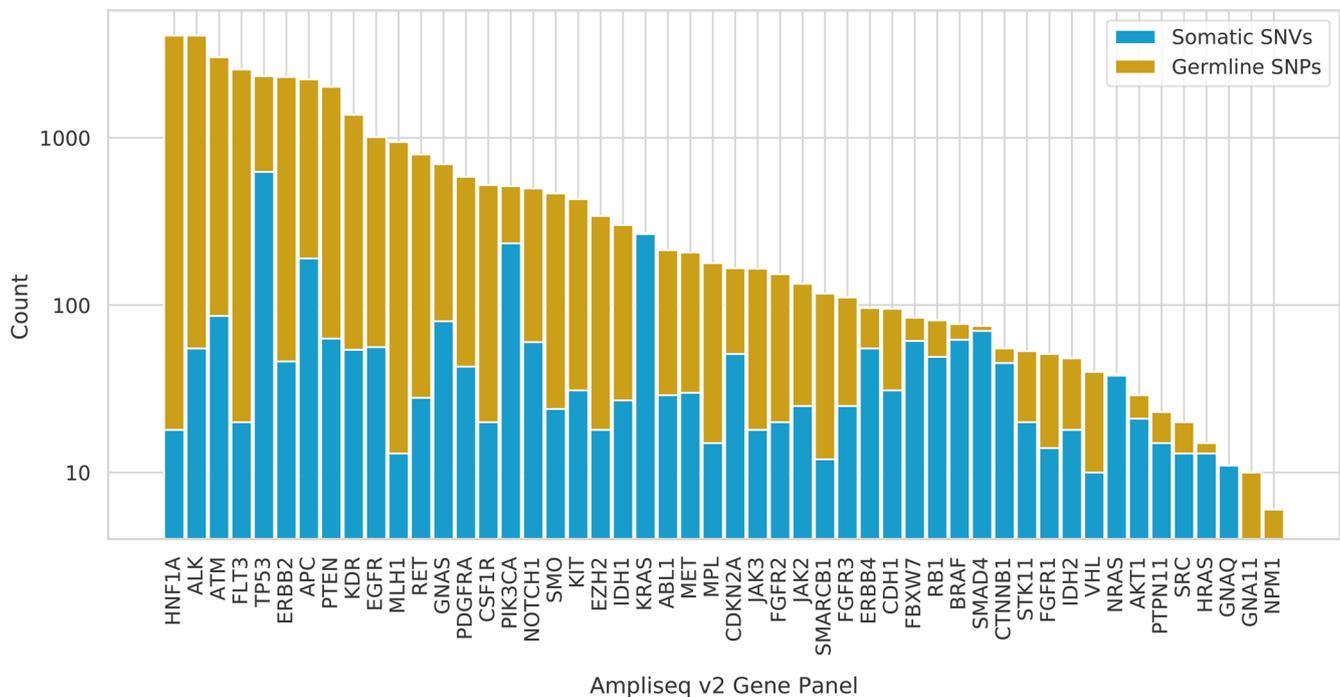
*Patient demographics.* This cohort comprised 1879 patients with multiple different tumor types (Supplemental Table 1; supplemental material available online with this article; <https://doi.org/10.1172/jci.insight.134824DS1>). All of these patients were used for tumor/normal NGS, that is, known tumor tissue NGS data were compared with those from normal nontumor tissue, which here was either same-patient blood or buccal (swab) tissue. The most common tumor samples ( $n > 100$ ) were from breast, colon, lung, soft tissue (including sarcoma), pancreatic, and ovarian cancer. The ages of all patients included in the study ranged from 0 years old for soft tissue sarcoma to 92 years old for bladder cancer. The lowest median age per tumor type was for soft tissue tumors, at 18 years old, and the highest median age was 72 years old for bladder cancer.

*SNVs in germline mutations compared with somatic mutations.* Somatic (tumor) and germline (normal tissue) DNA sequence information, specifically SNVs, were compared for all gene regions captured by whole-exome sequencing (WES) in 1879 patients. As shown in Figure 1, 92% of sites that do not comport with the hg19 reference human genome within tumor samples were eliminated as somatic events by comparison to the patient-matched germline genome; thus only approximately 8% of the SNVs identified in tumor genomes were tumor specific (not found in normal tissue). This highlights the importance of the tumor/normal sequence comparison we performed here (see Methods, *NGS-WES and somatic-specific variant calling*) (although many laboratories have bioinformatics and/or data analytic tools in place to mitigate this issue). In addition, 33.8% of germline single nucleotide polymorphisms (SNPs) within the gene panel were rare (not found after filtering through variant information domains, such as dbSNP; <https://www.ncbi.nlm.nih.gov/snp/>) and at risk of being falsely reported as somatic; at least 1 variant potentially falsely identified as somatic (i.e., rare or unannotated nonsomatic SNP) was found in 52% of patients. The genes with the highest number of germline rare SNPs were *ALK*, *HNFI1A*, *ATM*, *FLT3*, *TP53*, *ERBB2*, and *APC*. The genes with the most true somatic mutations were *TP53*, *PIK3CA*, *KRAS*, and *APC*.

*Discordance between DNA alterations and RNA expression of SNVs.* To determine rates of variant nonexpression in driver genes, 1417 patients with paired DNA- and RNA-Seq information were analyzed. The genes examined were based on the 50-gene Ampliseq cancer hotspot v2 panel (19) (see Methods).

Among 1417 patients, 2190 nonsynonymous somatic hotspot gene variants were found in DNA and 334 were not expressed in RNA (15.4%) as presented in Tables 1 and 2 and Figure 2. Of these 334 nonexpressed variants, 300 were from patients with common cancer types, which allowed tissue-specific gene expression levels to be ascertained. A high percentage — 85.7% (257 of 300) — of these variants were in genes that are typically expressed; thus, the rate of unexpected nonexpression of variants was estimated at 13.1%. At least 1 nonexpressed variant was found for 23% of patients. The genes with the highest discordance between DNA alteration and presence in RNA transcripts were *ALK*, *CSF1R*, *ERBB4*, *FLT3*, *GNAS*, *HNFI1A*, *KDR*, *PDGFRA*, *RET*, and *SMO*. The genes in which alterations were virtually always expressed at the RNA levels were *TP53*, *PIK3CA*, and *KRAS* (Figure 2).

The percentage of SNVs expressed as transcripts ranged from 58% in melanoma to 100%, the latter mainly in underrepresented tumor types, which may have prevented accurate frequency analysis (Table 1). Taken together, approximately 85% of SNVs were expressed at the RNA level. Excluding tumor types with



**Figure 1. Frequency of somatic single nucleotide variants and germline single nucleotide polymorphisms among genes in the Ampliseq panel.** Overall, 92% of single nucleotide variants (SNVs) identified from sequencing tumor genomes alone were of germline origin. (Ampliseq is a hotspot gene panel of the 50 most frequently altered genes in all cancers.) Interestingly, many of the genes with SNVs with the highest somatic specificity – *TP53*, *KRAS*, and *PIK3CA* – were also most likely to be expressed (see Figure 2). SNPs, single nucleotide polymorphisms.

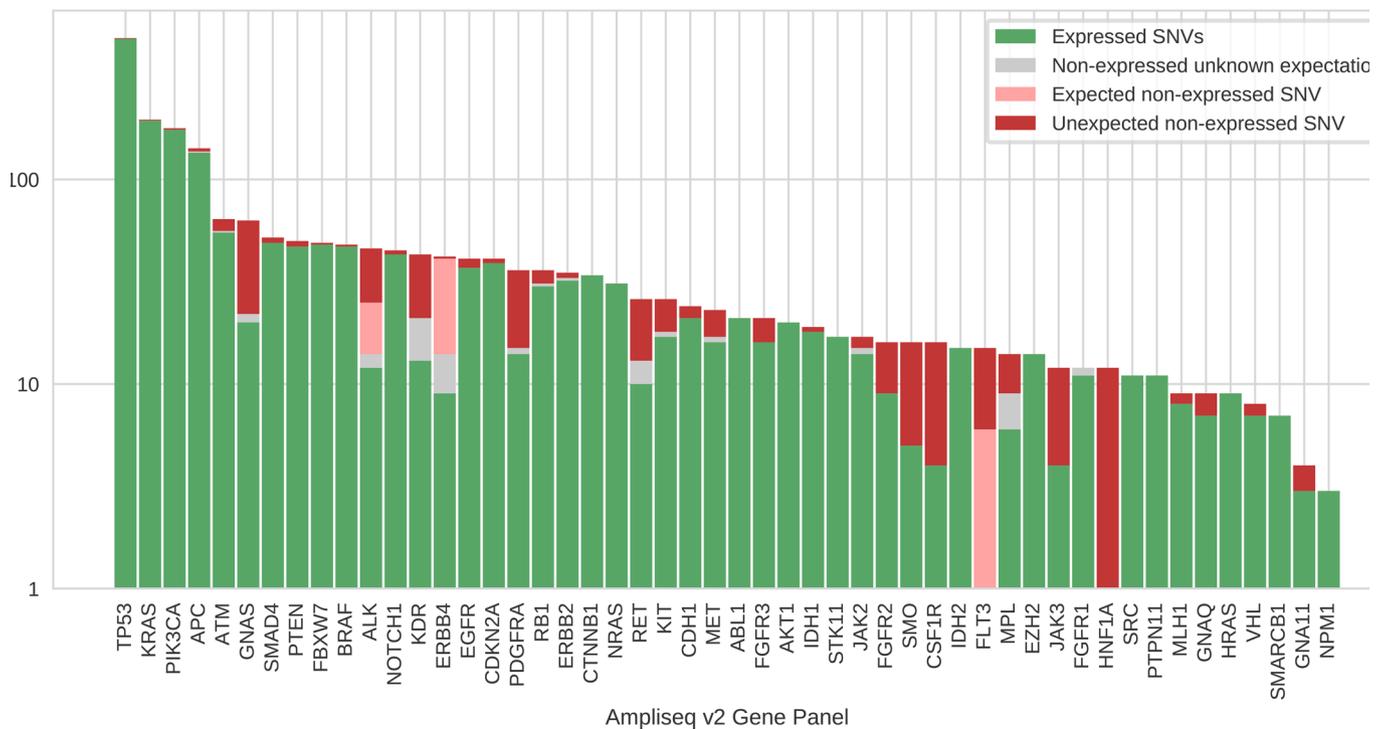
less than 5 SNVs (due to small numbers of patients), the cancers with the highest percentage of SNVs expressed at the RNA level were cervical, rectal, brain, ovarian, breast, and gastric (stomach) (ranging from 90%–96% expressed). The cancers with the lowest level of SNVs expressed at the RNA level were melanoma, nonmelanoma skin, and gallbladder cancer, ranging from 58%–69% expressed.

*Nonsynonymous mutational load and relationship to silenced variants.* Nonsynonymous mutation load (Table 1) was calculated based on WES, as described in the Methods. Nonmelanoma skin cancer and melanoma had the highest average nonsynonymous (amino acid–altering) mutational load. Mesothelioma and bile duct (extrahepatic tumors) had the lowest average nonsynonymous mutational load. There was a weak correlation (but statistically significant) relationship between mutational load and percentage of nonexpressed variants (Pearson  $R = 0.17$ ,  $P = 2.5 \times 10^{-7}$ ; Supplemental Figure 1).

## Discussion

In this report, we demonstrate discordance between genomic sequencing and transcriptome analysis for putative driver alterations (20), with the specific goal of revealing the frequency of variant non-expression in this cohort. This discordance, when it results in the lack of expression of a therapeutic protein target, may reflect a mechanism of resistance to therapy that has previously been underrecognized and should be subject to further investigation. It should be noted here that the samples analyzed were largely from primary tumors and were not compared with metastases, if available. This raises the possibility that genes that are silenced in the primary tumor may be expressed in a metastasis, making such metastases target for drugs not expected to be efficacious for the primary. In addition, while sequencing tumor genomes identifies somatic SNVs — some of which are targetable with FDA-approved drugs — our analysis here revealed a very high rate of germline mutations that could erroneously be reported as somatic in the absence of advanced bioinformatics algorithms. Interestingly, most *HNF1A* mutations were germline (and, importantly, all were silenced); both phenomenon would have significant implications for targeting this gene alteration.

Importantly, 23% of 1417 tumors had at least 1 somatic variant that was not expressed. Taking normal expression patterns into account, at least 13% of variants detected in DNA were unexpectedly not



**Figure 2. Expressed versus nonexpressed somatic single nucleotide variants among genes in the Ampliseq panel.** The proportion of somatic-specific single nucleotide variants (SNVs) for which transcripts were identified by RNA-Seq was determined. Tissue-specific gene expression patterns were established to ensure nonexpression of variants was not due to normal differential expression of genes (see Methods) and hence to differentiate expected versus unexpected nonexpression. At least 23% of patients had at least 1 somatic SNV that was not expressed. *TP53*, *KRAS*, and *PIK3CA* SNVs were almost always expressed at the RNA level, whereas *ALK*, *CSF1R*, *ERBB4*, *FLT3*, *GNAS*, *HNF1A*, *KDR*, *PDGFRA*, *RET*, and *SMO* and several other SNVs were often not expressed (see also Table 2). Since Ampliseq is a hotspot gene panel of the 50 most frequently altered genes in all cancers, nonexpressed SNVs may be clinically relevant.

expressed. We found a weak but statistically significant relationship between mutational load and the frequency of nonexpressed variants (Table 1 and Supplemental Figure 1). The relationship between mutational load and gene silencing, while weak, is an interesting finding that may merit further study. For example, nonmelanoma skin cancer had an exomic mutational load of 2660 and 33% silenced variants, while pancreatic cancer had a mutational load of 70.54 and 11% silenced variants. The genes with the highest discordance between DNA alteration and transcription were *ALK*, *CSF1R*, *ERBB4*, *FLT3*, *GNAS*, *HNF1A*, *KDR*, *PDGFRA*, *RET*, and *SMO*, while the genes in which alterations were virtually always expressed at the level of transcription were *TP53*, *PIK3CA*, and *KRAS*. In future studies, tumor type and other biological features could be considered along with mutational burden and expression.

While we did not determine the exact mechanisms involved in what appears to be transcriptional silencing, alteration of gene expression is characteristic of cancer cells; recently, there has been focus on the role of posttranscriptional editing and its effect in disease paradigms, particularly cancer, and the important context of the activity of silenced variants (21–24). Described mechanisms of RNA silencing on a global scale may involve miRNA, RNAi, and siRNA (25); may result in degradation of transcripts or silencing of transcription; and may be implicated in the gene SNV/transcript discordance seen here. For example, it is thought that mRNA sequences are cleaved by homologous siRNAs, which leads to RNAi and consequently silences expression (25). miRNAs can rapidly degrade transcripts (as well as inhibit translation), and the targets of miRNA are generally the genes and proteins previously implicated in drug resistance but even include other miRNAs involved in oncogenesis (12, 26).

Another key finding of our study was that 92% of potential SNVs from genomic sequencing were of germline origin (Figure 1) and that 33.8% of germline SNPs within the gene panel were rare and at risk of being falsely reported as somatic (even after filtering through databases for variant information, such as dbSNP (<https://www.ncbi.nlm.nih.gov/snp/>)); at least 1 such variant potentially falsely identified as somatic (i.e., nonsomatic variant) was found in 52% of patients (Supplemental Figure 2). However, it

**Table 1. Somatic SNVs in gene panel per cancer type, sorted by their respective expression percentages<sup>A</sup>**

Cancer type	No. mutations <sup>A</sup>	Expressed mutations	% Expressed (expressed mutations/no. of mutations)	Average nonsynonymous mutation load (across entire exome) <sup>B</sup>
Melanoma	76	44	58%	546.54
Skin (nonmelanoma)	67	45	67%	2660.33
Gallbladder	13	9	69%	77.45
Lymphoma	4	3	75%	134.92
Kidney	12	9	75%	344.61
Leukemia	4	3	75%	115.75
Soft tissue	8	6	75%	116.45
Head and neck	28	22	79%	120.71
Cancer type unknown	149	117	79%	437.24
Biliary tract (intrahepatic)	5	4	80%	71.9
Vaginal	5	4	80%	498.33
Thymic	5	4	80%	155.82
Bile duct (extrahepatic)	5	4	80%	59.6
Esophageal	45	37	82%	169.6
Mesothelioma	6	5	83%	52.25
Prostate	19	16	84%	95.18
Bone and soft tissue sarcomas	56	47	84%	123.81
Liver	19	16	84%	160.22
Other cancer	47	40	85%	105.82
Bladder	48	41	85%	235.7
Unknown primary	13	11	85%	160
Lung	168	143	85%	261.6
Anal	7	6	86%	250.4
Colon	332	289	87%	270.22
Oral/throat/thyroid cancers	40	35	88%	146.31
Uterine (endometrial)	28	25	89%	180.76
Pancreatic	118	105	89%	70.54
Gastric (stomach)	29	26	90%	136.58
Breast	225	202	90%	124.83
Ovarian	52	48	%	100.72
Brain	59	56	95%	90.61
Rectal	43	41	95%	241.29
Cervical	26	25	96%	229.38
Adrenal	2	2	100%	85.62
Small intestine	2	2	100%	241
Testicular	1	1	100%	393.17
Renal pelvis and ureter cancers	2	2	100%	125.8
Myeloma	1	1	100%	146
Penile	4	4	100%	476
Ampulla of vater	2	2	100%	108
Total	1775	1502	84.6%	

<sup>A</sup>Tumors with no SNVs were not counted. Mutations were those in the Ampliseq cancer hotspot v2 panel of 50 well-characterized driver genes (19).

<sup>B</sup>Nonsynonymous load refers to the raw counts of total nonsynonymous mutations across the entire exome; the “average” is the mean across the tumor type noted. SNV, single nucleotide variant.

should be kept in mind that many laboratories have sophisticated bioinformatics tools in place that can diminish the risk of reporting these germline variants as somatic alterations. Some of the genes most likely to have true somatic mutations were *TP53*, *PIK3CA*, and *KRAS*, all of which have well-defined roles in the pathogenesis and evolution of tumors across a multitude of tumor types.

This study has several important limitations. First, the data were derived retrospectively. Second, the sequencing information was curated from a deidentified database, and, hence, clinical correlations with outcome after cognate therapies, taking into consideration RNA expression, were not available. Such a study would be an important next step.

Table 2. Expressed and nonexpressed variants<sup>A</sup>

	Expressed SNVs	Unknown nonexpressed SNVs	Expected nonexpressed SNVs	Unexpected nonexpressed SNVs	% Unexpected nonexpressed SNVs
<i>HNF1A</i>	0	1	0	11	100.0
<i>CSF1R</i>	4	0	0	12	75.0
<i>SMO</i>	5	0	0	11	68.8
<i>GNAS</i>	20	2	0	41	67.2
<i>JAK3</i>	4	0	0	8	66.7
<i>KDR</i>	13	8	0	22	62.9
<i>PDGFRA</i>	14	1	0	21	60.0
<i>FLT3</i>	1	0	5	9	60.0
<i>RET</i>	10	3	0	13	56.5
<i>ALK</i>	12	2	11	21	47.7
<i>MPL</i>	6	3	0	5	45.5
<i>FGFR2</i>	9	0	0	7	43.8
<i>KIT</i>	17	1	0	8	32.0
<i>MET</i>	16	1	0	6	27.3
<i>GNA11</i>	3	0	0	1	25.0
<i>FGFR3</i>	16	0	0	5	23.8
<i>GNAQ</i>	7	0	0	2	22.2
<i>RB1</i>	30	1	0	5	14.3
<i>ATM</i>	55	1	0	8	12.7
<i>JAK2</i>	14	1	0	2	12.5
<i>VHL</i>	7	0	0	1	12.5
<i>CDH1</i>	21	0	0	3	12.5
<i>MLH1</i>	8	0	0	1	11.1
<i>EGFR</i>	37	0	0	4	9.8
<i>PTEN</i>	47	0	0	3	6.0
<i>ERBB2</i>	32	1	0	2	5.9
<i>SMAD4</i>	49	0	0	3	5.8
<i>IDH1</i>	18	0	0	1	5.3
<i>CDKN2A</i>	39	0	0	2	4.9
<i>NOTCH1</i>	43	0	0	2	4.4
<i>APC</i>	135	2	0	5	3.6
<i>ERBB4</i>	9	5	27	1	2.7
<i>BRAF</i>	47	0	0	1	2.1
<i>FBXW7</i>	48	0	0	1	2.0
<i>PIK3CA</i>	175	0	0	3	1.7
<i>KRAS</i>	194	0	0	2	1.0
<i>TP53</i>	487	0	0	4	0.8
<i>HRAS</i>	9	0	0	0	0.0
<i>SRC</i>	11	0	0	0	0.0
<i>PTPN11</i>	11	0	0	0	0.0
<i>SMARCB1</i>	7	0	0	0	0.0
<i>ABL1</i>	21	0	0	0	0.0
<i>FGFR1</i>	11	1	0	0	0.0
<i>EZH2</i>	14	0	0	0	0.0
<i>IDH2</i>	15	0	0	0	0.0
<i>STK11</i>	17	0	0	0	0.0
<i>AKT1</i>	20	0	0	0	0.0
<i>NRAS</i>	31	0	0	0	0.0
<i>CTNNB1</i>	34	0	0	0	0.0
<i>NPM1</i>	3	0	0	0	0.0

<sup>A</sup>Tissue-specific gene expression patterns were established to ensure nonexpression of variants was not due to normal differential expression of genes (see Methods) and hence to differentiate expected versus unexpected nonexpression. See also Figure 2. SNV, single nucleotide variant.

There have been several meta-analyses assessing the association and effect of biomarker-driven clinical trials and patient outcomes in terms of response rates as well as progression-free and overall survival. In meta-analyses of over 80,000 patients in phase I, II, and III trials, the subsets of patients treated with personalized biomarker-driven therapies had significant improvements in all outcome parameters (27–29), with genomic biomarkers producing response rates over 40% versus approximately 5% for no biomarkers. However, these numbers still leave many patients without response or with fleeting responses, even when matched by genomic biomarkers. A myriad of resistance mechanisms may account for unresponsiveness. They include, but are not limited to, driver coalterations in other genes, new mutations that emerge, and inadequate drug engagement with the target. However, our current observations indicate that a significant fraction of somatic variants are not expressed at the RNA level and that almost one-quarter of cancers harbor at least one unexpressed somatic variant (Figure 2 and Table 2). Since the mutations explored were those in the Ampliseq cancer hotspot v2 panel of 50 well-characterized driver genes, these nonexpressed variants are likely to be clinically relevant. Perhaps not coincidentally, strong drivers such as *TP53*, *PIK3CA*, and *KRAS* were most likely to have true somatic mutations and were almost always expressed. In future studies, we will leverage paired DNA-Seq/RNA-Seq analyses to further explore other relevant and translatable findings, including neoantigen expression, novel fusion transcript detection, indel analysis, and alteration in copy number.

Almost one-quarter of 1417 tumors had a potentially deleterious gene mutation that was not expressed at the RNA level. The effect of gene silencing on response versus resistance after targeted therapies matched to oncogenic DNA alterations requires large-scale clinical correlative studies. Gene silencing may also be relevant to immunotherapy, as silenced alterations would not be expected to produce presented neoantigens; this too merits additional study. Taken together, our observations suggest that proper pharmacologic prosecution of a malignancy likely requires a full molecule portrait that includes transcriptomic interrogation of expressed versus silenced genomic alterations.

## Methods

*Sample preparation and data collection.* Germline (normal tissue) and tumor NGS were performed in tandem using FFPE slides or tissue blocks and matched blood or buccal swab samples from 1879 unselected patients with advanced cancer (Clinical Laboratory Improvement Amendments, NantHealth; <https://nanthealth.com>). Up to ten 5- $\mu$ m tissue sections were cut from blocks using standard microtomy techniques. The initial tissue section was subjected to H&E staining for histological analysis. An internal board-certified pathologist marked the regions of the malignancy on image of H&E-stained slides and confirmed the histological diagnosis. Using the marked-up image as a guide, tissue sections were macrodissected to collect the desired tumor regions. Each collected cell population was then extracted using the QIAGEN QIAamp DNA FFPE Tissue Kit for DNA. Extracted material was then quantified using a Qubit fluorometer.

*NGS-WES and somatic-specific variant calling.* Somatic-specific SNVs were identified as previously described (30) by WES of tumor and normal DNA. In brief, NGS libraries were captured to exome regions using xGen Exome Research Panel v1.0 (IDT), and libraries were prepared using the KAPA Hyper prep kit (Kapa Biosystems). DNA libraries were sequenced to a target depth of  $\times 200$  for tumor sample and  $\times 100$  for normal samples on the Illumina HiSeq platform. In order to obtain position-specific read support, raw reads were aligned to the hg19 reference genome using the Burrows-Wheeler Aligner (<http://bio-bwa.sourceforge.net>) and converted to BAM format using SAMtools software (<http://www.htslib.org>). To remove support bias from polymerase chain reaction effects, duplicate reads were removed using SAMBLASTER (<https://github.com/GregoryFaust/samblaster/commit/a5274027e4c887ef6ab14f3226c0e36beb4f4f25>). Any site with  $\geq 8$  nonduplicate reads, mapping quality  $\geq 20$  (according to SAMtools), and a nonreference allele was considered as a possible SNV. If the tumor and normal genotypes were identical, then the variant was classified as germline and removed from the pool of possible SNVs. Instances where the normal genotype was heterozygous and the tumor genotype was homozygous suggest regions of loss of heterozygosity, as opposed to new SNV mutations, and such variants were so classified.

The remaining potential SNVs were filtered to those with high confidence using several quality metrics (see North et al. for details, ref. 30), including coverage metrics as well as those derived from SomaticSniper v1.0.5.0 (<http://gmt.genome.wustl.edu/packages/somatic-sniper/>), a Bayesian model that reports the probability of the observed tumor genotype given the prior probability of the reference genotype and the fraction of heterozygous positions in the human genome. A single variant cell format file was produced for each tumor sample describing somatic-specific SNVs.

*Calculation of potential for falsely identifying SNV as somatic.* Germline variants that appear in less than 0.1% of the general population may be falsely reported as somatic if analysis lacks a complete germline comparator (i.e., lacking matched tumor/normal sequencing). In order to calculate the rate of potentially falsely identifying a variant as somatic by performing tumor-only sequencing, the population allele frequency (PAF) for each potential variant was found in a public database of >125,000 exomes (gnomAD; <https://gnomad.broadinstitute.org>), and those with PAF <0.1% were counted as potential false positives for being somatic specific.

*Nonsynonymous mutational load determination.* Mutational load was directly observed in whole-exome sequences by counting all tumor-specific nonsynonymous exonic SNVs.

*RNA-Seq and expressed variant calling.* RNA extraction was performed on samples with sufficient tissue remaining after NGS using an RNeasy FFPE Kit (QIAGEN). Extracted material was then quantified using a Qubit fluorometer (Thermo Fisher Scientific). RNA-Seq libraries were prepared using the KAPA Stranded RNA-Seq Kit with RiboErase (Kapa Biosystems) and sequenced to a target depth of 200 M reads on the Illumina HiSeq platform (Illumina). RNA samples were aligned to RefSeq build 73 transcriptome using Bowtie2 v2.2.6 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>). Approximately 14% of samples had a low alignment mapping rate due to RNA degradation or contamination and were removed to ensure quality. In total, excluding samples with insufficient tissue and those with poor alignment, 1417 of the original 1879 samples were included in this analysis cohort. To determine if a given SNV identified in NGS was expressed as mRNA, an approach extremely similar to published tools, such as RADIA (15), was used: RNA-Seq mapped reads spanning the variant site were counted, and if more than 2 RNA-Seq reads containing the SNV were identified, the variant was considered to be expressed.

To reduce confounding by passenger mutations that accrue stochastically (especially in large genes such as *TTN*), analysis was focused on rates of variant expression only within putative driver genes. The Ampliseq cancer hotspot v2 panel was selected as a representative set of 50 well-characterized driver genes (19). Counts of expressed somatic-specific variants anywhere within these 50 genes (not just hotspot mutations) are reported here.

Tissue-specific gene expression patterns were established to ensure nonexpression of variants was not due to normal differential expression of genes as follows: tissue categories with at least 10 patient samples were used to establish minimum gene expression for all 50 hotspot genes; and, for each gene and tissue combination, if the minimum gene-level RSEM transcripts per million within the tissue- category was 0, then patients with nonexpressed variants within that gene and tissue combination were removed from the total of nonexpressed variants. Rarer tissue types were excluded from analysis.

*Study approval.* All data were collected and analyzed in a fashion consistent with applicable law for deidentified databases (and considered Internal Review Board exempt by University of California, San Diego).

## Author contributions

CWS, SKR, SK, and RK developed the concept; SK, RP, CWS, JZS, CJV, SCB, and SKR collected and/or analyzed data; JJA and RK wrote and revised manuscript. All authors approved final version.

## Acknowledgments

This study was funded in part by NantHealth, the Joan and Irwin Jacobs Fund, and National Cancer Institute grant P30 CA023100 (to RK). Many thanks to Patricia Spilman for editing help.

Address correspondence to: Razelle Kurzrock, Center for Personalized Cancer Therapy, University of California, San Diego, Moores Cancer Center, 3855 Health Sciences Drive, La Jolla, California 92093, USA. Phone: 858.822.3050; Email: [rkurzrock@ucsd.edu](mailto:rkurzrock@ucsd.edu).

1. Bieg-Bourne CC, et al. Next-generation sequencing in the clinical setting clarifies patient characteristics and potential actionability. *Cancer Res.* 2017;77(22):6313–6320.
2. [No authors listed]. 2018 Decision Memo for Next Generation Sequencing (NGS) for Medicare Beneficiaries with Advanced Cancer (CAG-00450N).Centers for Medicare & Medicaid Services. <https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=290>. Accessed May 7, 2020.
3. Subbiah V, Kurzrock R. Universal genomic testing needed to win the war against cancer: genomics IS the diagnosis. *JAMA Oncol.* 2016;2(6):719–720.
4. McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell.* 2017;168(4):613–628.
5. Li ZW, Dalton WS. Tumor microenvironment and drug resistance in hematologic malignancies. *Blood Rev.* 2006;20(6):333–342.

6. Higgins CF. ABC transporters: from microorganisms to man. *Annu Rev Cell Biol.* 1992;8:67–113.
7. Mansoori B, Mohammadi A, Davudian S, Shirjang S, Baradaran B. The different mechanisms of cancer drug resistance: a brief review. *Adv Pharm Bull.* 2017;7(3):339–348.
8. Matsui A, Ihara T, Suda H, Mikami H, Semba K. Gene amplification: mechanisms and involvement in cancer. *Biomol Concepts.* 2013;4(6):567–582.
9. Barouch-Bentov R, Sauer K. Mechanisms of drug resistance in kinases. *Expert Opin Investig Drugs.* 2011;20(2):153–208.
10. Incorvaia L, et al. “Back to a false normality”: new intriguing mechanisms of resistance to PARP inhibitors. *Oncotarget.* 2017;8(14):23891–23904.
11. Wu Q, et al. Methylation of miR-129-5p CpG island modulates multi-drug resistance in gastric cancer by targeting ABC transporters. *Oncotarget.* 2014;5(22):11552–11563.
12. Liu H, Wu X, Huang J, Peng J, Guo L. miR-7 modulates chemoresistance of small cell lung cancer by repressing MRP1/ABCC1. *Int J Exp Pathol.* 2015;96(4):240–247.
13. Jewer M, Findlay SD, Postovit LM. Post-transcriptional regulation in cancer progression: Microenvironmental control of alternative splicing and translation. *J Cell Commun Signal.* 2012;6(4):233–248.
14. Kechavarzi B, Janga SC. Dissecting the expression landscape of RNA-binding proteins in human cancers. *Genome Biol.* 2014;15(1):R14.
15. Radenbaugh AJ, et al. RADIA: RNA and DNA integrated analysis for somatic mutation detection. *PLoS One.* 2014;9(11):e111516.
16. Wilkerson MD, et al. Integrated RNA and DNA sequencing improves mutation detection in low purity tumors. *Nucleic Acids Res.* 2014;42(13):e107.
17. Coudray A, Battenhouse AM, Bucher P, Iyer VR. Detection and benchmarking of somatic mutations in cancer genomes using RNA-seq data. *PeerJ.* 2018;6:e5362.
18. Yizhak K, et al. A comprehensive analysis of RNA sequences reveals macroscopic somatic clonal expansion across normal tissues. bioRxiv. <https://www.biorxiv.org/content/10.1101/416339v1>. Published September 13, 2018. Accessed May 7, 2020.
19. Thermo Fisher Scientific. Ion AmpliSeq Cancer Hotspot Panel v2. <https://assets.thermofisher.com/TFS-Assets/LSG/brochures/Ion-AmpliSeq-Cancer-Hotspot-Panel-Flyer.pdf>. Accessed May 14, 2020.
20. Kurzrock R, et al. Seventeen percent of NGS 50 gene panel variants are not expressed in RNAseq. *J Clin Oncol.* 2018;36(suppl\_15):12118.
21. Han ZJ, Feng YH, Gu BH, Li YM, Chen H. The post-translational modification, SUMOylation, and cancer (Review). *Int J Oncol.* 2018;52(4):1081–1094.
22. Lian H, Wang QH, Zhu CB, Ma J, Jin WL. Deciphering the epitranscriptome in cancer. *Trends Cancer.* 2018;4(3):207–221.
23. McDuff FK, Lim SV, Dalbay M, Turner SD. Assessment of the transforming potential of novel anaplastic lymphoma kinase point mutants. *Mol Carcinog.* 2013;52(1):79–83.
24. Schönherr C, et al. The neuroblastoma ALK(I1250T) mutation is a kinase-dead RTK in vitro and in vivo. *Transl Oncol.* 2011;4(4):258–265.
25. Agrawal N, Dasaradhi PV, Mohammed A, Malhotra P, Bhatnagar RK, Mukherjee SK. RNA interference: biology, mechanism, and applications. *Microbiol Mol Biol Rev.* 2003;67(4):657–685.
26. Sun M, Estrov Z, Ji Y, Coombes KR, Harris DH, Kurzrock R. Curcumin (diferuloylmethane) alters the expression profiles of microRNAs in human pancreatic cancer cells. *Mol Cancer Ther.* 2008;7(3):464–473.
27. Schwaederle M, et al. Association of biomarker-based treatment strategies with response rates and progression-free survival in refractory malignant neoplasms: a meta-analysis. *JAMA Oncol.* 2016;2(11):1452–1459.
28. Schwaederle M, et al. Impact of precision medicine in diverse cancers: a meta-analysis of phase II clinical trials. *J Clin Oncol.* 2015;33(32):3817–3825.
29. Jardim DL, et al. Impact of a biomarker-based strategy on oncology drug development: a meta-analysis of clinical trials leading to FDA approval. *J Natl Cancer Inst.* 2015;107(11):djv253.
30. North JP, et al. Cell of origin and mutation pattern define three clinically distinct classes of sebaceous carcinoma. *Nat Commun.* 2018;9(1):1894.